Comments from Jeremy Buck, US Fish and Wildlife Service, Portland, Oregon on the *Draft Work Plan with Quality Assurance Project Plan for Smallmouth Bass Acoustic Telemetry and Tissue Sampling and Crayfish Tissue Sampling at River Operable Unit, Bradford Island*
July 9, 2020

The following description of the data quality objective (DQO) process primarily focuses on key items of importance documented in U.S. Environmental Protection Agency (EPA) guidance for data quality, and serves as a few good reminders to consider when developing a sampling plan.

When using a classical statistics approach to guide decisions, as is the proposed approach in the *Draft Work Plan*, there are a few requirements needed for each Contaminant of Potential Concern (COPC) within each decision unit to ensure data representativeness. Data representativeness means the degree to which a given sample or samples can be used to estimate the property of the area (volume) that is it supposed to represent within acceptable limits. These requirements include a:

1. Reasonably accurate estimate of Probability Density Function (Histogram) based on previous data
2. Reasonably accurate estimate of the standard deviation based on previous data
3. Correct selection of an appropriate statistical test method such as an equation for calculating an Upper Confidence Level (UCL) based on the Probability Density Function
4. Correct selection of an appropriate statistical sample size method (equation) base on the Probability Density Function or histogram

For decision making purposes, there are a number of pitfalls to be cautious of when developing a sampling plan. A good sampling design with good data quality objectives should prevent the following:

- Failure to define population accurately
- Failure to collect representative samples from the population of interest
- Failure to obtain representative data from the population of interest
- Failure to accurately determine the frequency distribution of the COPCs
- Failure to accurately determine the standard deviation of the COPCs
- Failure to select the appropriate statistical method for generating adequate samples
- Failure to use the appropriate UCL in making the decision

The list of project quality objectives in the current Table 4 of the *Draft Work Plan* are very general, and more details are needed within each DQO step in order to resolve concerns at the site. The DQOs listed below in Table 1 are proposed in order to prevent the statistical pitfalls described above, and to promote discussion that will lead to adding specific details to the table.

Specific statistical parameters can be estimated from previously collected data from the site, or decided upon in discussion with the TAG and other professionals. The process for developing DQOs (as presented directly below) and summarized in Table 1 are slightly different than what is currently listed in Table 4 of the *Draft Work Plan*. The process listed below follows a standardized format recommended from the following resources:

U.S. Environmental Protection Agency. 2000. Data quality objectives process for hazardous waste site investigations. EPA QA/G-4HW. Office of Environmental Information, Washington, DC.

With supporting information from:

Deming, W.E. 1950. Some theory of sampling. Dover publications, New York. 640 pp.

Tindall, S. 2010. Managing uncertainty with systematic planning: Developing defensible sample designs for environmental decision making. Training class manual. QE3C, Inc. May 15-17, 2010, Portland, Oregon.

U.S. Environmental Protection Agency. 2000c. Guidance for data quality assessment: Practical methods for data analysis. EPA QA/G-9, QA00 Update. Final Report EPA/600/R-96/084. Report EPA/600/R-96/084.

The basic principles for DQOs as outlined in the above documents and that are summarized in Table 1 are presented below. Comments in red text indicated issues that still need to be addressed in order to produce a sampling plan that will appropriately resolve the remaining questions at the site.

1) *Problem Statement*: The **input** for this step is the systematic planning and scoping process, information from stakeholders and interviews with people knowledgeable about the site, identification of the contaminants of potential concern (COPCs), and development of the conceptual site Model (CSM). The CSM is used to constrain the problem statement. The objective is to clearly define the problem so that the focus of the project will be unambiguous. The **output** of this step is a concise problem statement describes the problem as it is currently understood and the conditions that are causing the problem. The individual components of this step include the following (many of which have been completed for the Bradford Island river operational unit):

    a. Identify COPCs [completed]
    b. Provide rationale for COPC exclusions [still need to agree on which COPCs can be excluded and which of the 209 PCB congeners we can focus on]
    c. Specify release mechanisms [complete, though how PCBs get into some tissues from which matrix is still a mystery]
    d. Identify fate and transport mechanisms [complete, though transport and some fate data still are unknown and it is unknown how far PCBs may have migrated in the Forebay–can decisions be made without it or is more necessary?]
    e. List potential receptors [complete]
    f. Estimate COPC distributions- Spatial & Frequency distributions which determines the number of samples [incomplete or not presented for crayfish, clams, bass]
    g. Discuss decision drivers [incomplete]
    h. Write CSM Summary Narrative (this will help determine the placement and importance of samples within the Bradford Isl., Forebay, Goose Isl., Ref areas

and refine decision units within those areas) [CSM graphic complete, CSM narrative might be helpful]

2) *Identify Decision*s: The objective is to develop decision statements that require environmental data to address the problem statement. The **input** for this step is information collected for deriving the problem statement. The **output** of this step is a decision statement derived from the Principle Study Questions (the questions that must be addressed in order to resolve the problem statement) and Alternative Actions (actions that may be taken as a result of answering the question). The individual components of this step include the following:

    a. Identify Principal Study Questions [complete but revisions suggested as indicated in table below]
    b. Define Alternative Actions [incomplete and is a critical component of this step, recommendations in table below]
    c. Define error if Alternative Actions incorrectly taken [incomplete, but helpful if defined]
    d. List consequences of errors [incomplete, but helpful to address]
    e. Rate severity of consequences [incomplete, but helpful]

3) *Identify Inputs*. The objective of this step is to identify applicable information and quality of information needed for making the decisions, and whether new data are required or if existing data are sufficient. The **input** of this step is the decision statement. The **output** is the information needed to resolve the decision statement. The individual components of this step include the following:

    a. Specify environmental variables to be measured [complete but some unclear]
    b. List general sources of information [complete]
    c. Determine whether the information exists [complete but unclear]
    d. Determine the general level of quality required for the data [incomplete or not presented for tissue data, needs review]
    e. Evaluate appropriateness of existing data through usability assessment [incomplete or not presented; this would be very necessary and would be very helpful for this process]
    f. Confirm appropriate measurement methods exist [complete but discussion needed]
    g. Specify the matrix to be measured [complete, more discussion needed on WHEN data should be gathered, such as waiting for SPME data to influence data collection]
    h. Identify the action level and basis for level [incomplete, other than comparing to reference values, needs discussion]
    i. Specify required detection limits [complete]
    j. Specify precision required [more review/discussion of this would be helpful]
    k. Specify the accuracy required [more review/discussion of this would be helpful]

4) *Specify Boundaries.* The objective of step 4 is to set the boundaries for decision making. It provides the biggest single opportunity for managing uncertainty. The **input** is the results from comprehensive scoping and professional judgment. Professional judgment is key in defining the kind and size of sampling units, delineating homogeneous and heterogeneous areas, and classifying sites into strata in ways that will reduce sampling error. It is important for finding distinct populations of interest and separating them for measurement. This ensures the data are representative, and decisions are made and final actions taken based on samples from a well-defined population. However, professional judgement is NOT allowed to influence the final selection of the particular locations of samples within the decision unit. Eliminating any professional judgement from influencing the final sample location selection will ensure that selection bias is eliminated and sampling tolerance will be measurable and controllable (Deming 1950). The **output** of this step is the final unit of decision-making. The individual components of this step include the following:

    a. Define the population of interest [there seems to be separate populations here based on previous data, as the Forebay and Goose Isl. samples are much different from north shore Bradford Island samples, and the reference populations need discussion. There are also curious outlies in the tissue data]
    b. Define the spatial boundaries of the decision statement [needs to be refined]
    c. Determine the temporal boundary of the problem [incomplete or not presented and needs discussion]
    d. Define the scale of decision making [need discussion on the basis for selecting decision units]
    e. Identify any practical constraints on data collection [complete but unclear what will be done when the constraint is encountered]

5) *Define Decision Rules.* The objective of step 5 is use the parameter of interest, the unit of decision making, the action level, and alternative actions to form decision rules. The **input** is the results from steps 1 through 4. The **output** of this step is the If/Then Decision Rule Statement(s). The individual components of this step include the following:

    a. Specify the parameter of interest [parameter estimate for the population needs to be refined]
    b. Confirm the Action Level [proposed as a reference level, yet the action level could be background, reference, or a risk value]
    c. Develop a Decision Rule [incomplete – decision rule is a "if...then..." statement that incorporates the parameter of interest, the unit of decision making, the action level, and the action(s) that would result from resolution of the decision]

6) *Specify Error Tolerances.* The objective of step 6 is to specify the tolerable limits on decision errors, which are used for limiting uncertainty in the data and to reduce the chance of making a decision error to a tolerable level. The two types of decision errors we are concerned with and want to limit are: cleaning up a clean site and walking away from a dirty site. The **input** is the decision rules from step 5. The **output** is the bounds

of the gray region and decision error tolerances. The gray region is a range of possible parameter values within which the consequences of a decision error are relatively minor. It is bounded on one side by the upper bound of the gray region (the action level), and on the other by the parameter value where the consequences of decision error begins to be significant (the lower bound of the gray region). The individual components of this step include the following:

a. Determine the variability of environmental variables [incomplete, needs assessment to determine appropriate number of samples]
b. Identify the decision errors [incomplete, needs more specificity and details]
c. Choose the null hypothesis [we need to concur that the null is that the site is dirty (contaminated) and we are trying to obtain sufficient data to disprove the null and support the alternative]
d. Specify the boundaries of the gray region [incomplete]
e. Assign probability limits on either side of the gray region [incomplete]

7) *Optimize Sample Design.* The objective of step 7 is to identify the most resource effective data collection and analysis design that satisfies the Planning Process Objectives specified in the preceding 6 steps. This requires proposing, comparing, and understanding sampling design alternatives for a specific project and then selecting the optimal design that meets the project objectives. The **input** is based on revisiting and modifying, as needed, decisions made in the previous 6 steps, and checking to see if number of samples or other information required for each alternative design exceeds project resource constraints. The **output** is the selection of the most optimal sample design. The emphasis here is that any alternative plan should incorporate probability sampling and not rely on judgmental sampling. The individual components of this step include the following:

a. Review Planning Process outputs from Steps 1-6 to be sure they are internally consistent [incomplete, needs more discussion]
b. Develop alternative sample designs [incomplete, needs discussion as to level of detail needed on alternative plans]
c. For each design option, select needed mathematical expressions [incomplete]
d. Select the optimal sample size that satisfies the Planning Process Objectives for each data collection design option [incomplete, discussion needed for level of detail necessary for this]
e. Check if the number of samples exceeds project resource constraints [incomplete]

Table 1. Proposed revisions to Data Quality Objectives

[HYPERLINK \l "_bookmark0"] **State the Problem**	**Step 2: Identify the Decision (see following table for decision matrix)**	**Step 3: Identify Information Inputs**	**Step 4: Define the Boundaries of the Study**	**Step 5: Define Decision Rules**	**Step 6: Specify Error Tolerances**	**Step 7: Optimize Sample Design**
In order to confirm that early remediation efforts at the Bradford Island river operational unit were successful in reducing concentrations of PCBs and other contaminants [or are no longer contributing to concentrations in aquatic organisms], current data regarding concentrations in tissue are needed. In order to understand if additional sources of PCBs occur within the river operational unit of Bradford Island, data regarding PCB concentrations in sedimentary organisms (clams and crayfish) and location data on mobile organisms (smallmouth bass) are needed. Note: I inserted some clam PSGs here as an example, but others listed in the clam QAPP may be added. More also could be added here for the movement telemetry data.	Determine whether the site contributes PCBs to bass or crayfish body burdens in excess of reference or action levels and requires further source identification and remediation; if not, then rely on bass or crayfish for long-term monitoring only. Determine whether PCB concentrations in bass have remained elevated at the site over time and require further identification of source materials; if not, then set up long-term monitoring or equivalency analysis using these tissues. Determine whether bass location data can help identify sources at the site that require remediation or follow up investigation using less mobile receptors; if not, use other evidence to establish where source materials are located. Determine whether crayfish consumed by bass contain higher concentrations than crayfish otherwise available at the site and help identify sources and delineate decision units; if not, use other lines of evidence to identify source areas.	PCB congener specific data- high quality HRGC/HRMS Aroclor PCBs – GC/ECD (need EPA lab method numbers here). Other constituents? High variability in previous tissue PCB data- (see box plots submitted by DEQ) needs further Data Quality Assessment to identify usability. Previous data may not be representative of the population but is reasonable for other CSM purposes. Action level for PCBs in approx. 20 to 100 µg/kg. Tissue matrices to be measured are crayfish, smallmouth bass, and clams (Corbicula). Detection limits are listed in QAPP. Precision <20%; Accuracy 75 to 125%.	The spatial boundaries within the site will be defined as decision units (DUs) based on previously sampled data (DUs still need to be refined). Sampled populations will occur within DUs in the north shoreline of Bradford Island, Forebay, Goose Island, possibly other reference area if one of the listed DUs does not suffice for reference. Targeted sampling locations will be selected based on a stratified random grid, with any location changes in the field based on random selection listed in field revision protocols. Whole body tissues will be analyzed (entire bass minus stomach contents which will be purged and archived, entire crayfish, clam minus shell (depurated?). A total of n crayfish make up one composite, and n composites will be collected within each DU. A total of n clams will make up one composite, and n composites will be collected within each decision unit. Define temporal boundary here- Fall 2020? Practical constraints- ESA permits, flow regime/dam constraints, substrate constraints and tissue samples not occurring at desired location. Scale of decision making- All possible tissue samples within each DU represented by x by x meter of surface area, collected during the fall. Each DU will be sampled for 40 discrete bass, n composites of crayfish, and n composites of clams.	The population parameter of interest will be the true mean as estimated by the one-sided 95% UCL. If the true mean (as estimated.by the 95% UCL calculated using the sample mean) total PCB concentration in tissues within each DU is $\geq$ the action level [100 µg/kg], then the DU is a source of contamination requiring further delineation; if not, PCB sources will be evaluated elsewhere. This can also be demonstrated using a one-sample t-test equation, where calculated t = (sample mean – Action Level) / (std. dev / sqrt(n)). If calculated t is less than table value, decide site is clean or not a contributing source. Note: 100 µg/kg is used as an example here for bass and will be discussed. Other values needed for clams and crayfish.	The variability of the environmental variable (COPC) will be evaluated using estimated standard deviations of each constituent for each tissue (from previous data on tissues within or near the DU, or by dividing the upper or lower range by 2 or 3). The number of samples required from each DU can be estimated as the square root of the standard deviation. The null hypothesis is that the site is contaminated (or each DU continues to act as a contributing source of contaminants (i.e., PCBs). The two types of decision errors are claiming a site or DU is a contributing source when it really isn't, or claiming it is not contributing when it really is. Which decision error has the most severe consequences near the action level? Setting Error tolerances: The alpha error is set to 5%. The beta error is set to 20%. The upper bound of the gray region is the action level (such as 100 µg/kg for bass) The lower bound of the gray region is ½ the action level or calculated based on PDF for total PCBs. This is the value where the consequences of the decision error begin to be significant.	Present alternative designs and determine which are the most cost effective. Different designs will consist of different numbers of samples or other statistical parameters which will may increase or decreases costs. The optimal design will be the least cost method that effectively balances decision errors to tolerable levels. I see this as a "sensitivity" analysis for optimal sample design.

Table 2. Principle Study Questions (PSQs) and Alternative Actions (AAs) matrix table.

PSQ #	PSQ	AA#	AA
1	Are tissue concentrations in bass or crayfish at Braford Island higher than the reference site and/or the action level?	1	Yes – further identify and remediate sources of contamination
		2	No – no further action for bass and crayfish other than for long term monitoring
2	Have tissue concentrations in bass or crayfish at the site increased or remain unchanged over time? Note: this may require a longer term equivalency-type analysis and the question could also be asked: Can variation in concentrations in bass or crayfish be characterized sufficiently to detect changes over time using an equivalency analysis?*	1	Yes – identify source of contamination
		2	No – set up long-term monitoring or equivalency analysis.
3	Do bass movements indicate where potential exposure to contaminated sediment is occurring at the site?	1	Yes – confirm contamination indicated from movements by sampling sediment or immobile receptors.
		2	No - use other evidence to establish where source materials are located.
4	Do crayfish consumed by bass contain higher concentrations than crayfish otherwise available to bass?	1	Yes – use bass location data to identify source areas and decision units.
		2	No – use other lines of evidence to identify source areas.
*Note that extreme variation in contaminant concentrations may preclude using classic statistical approaches to answer the PSQ, or may indicate improper selection of decision units (i.e., decision units are too heterogeneous or too inclusive of multiple populations of interest, and the populations should be sampled as distinct populations and considered separately). Populations could be separated and identified separately by looking at previously collected data and decreasing decision unit size to incorporate (potentially) a smaller degree of variation.			